Incidental Changes to Training Capabilities Due to Technological Improvements

Justin H. Prost, Ph.D.[1] and Brian Schreiber, M.S.[2]
Lumir Research Institute, Tempe, Arizona, 85282

and

Winston Bennett, Jr. Ph.D.[3]
Air Force Research Lab, Mesa, Arizona, 85212

During the development and enhancement of simulator systems, new technologies are incorporated to improve the training capabilities of the system. The degree to which changes to the simulator system improve training capabilities in anticipated areas is of primary interest; however, an evaluation of benefits from newly incorporated technologies should consider the impact on training capabilities in other, less directly related areas. The current study utilized previously developed methodology which identifies and evaluates the fidelity and deficiencies of a simulation system. The process involved utilizing Mission Essential Competencies and Subject Matter Experts to determine the training capabilities of a simulation system. The work was applied to the development of a Deployable Tactical Trainer as a method of providing feedback on the system. Data from the process were obtained throughout the development of the simulator, and the current study presents a comparison of the evaluation of the system before and after a meaningful upgrade to the technology of the simulation system. At issue is determining whether or not the upgrade resulted in the expected changes to the training capabilities. Incidental changes to training capabilities of the system were identified and are discussed, presenting an overall evaluation of the expected versus unexpected effects on the training capabilities of the system. Implications for the importance of building a global awareness of the simulator system are discussed. The growing dependence on simulation systems for training increases the impact of changes to the system on pilot readiness.

I. Introduction

SIMULATOR training environments are an effective training supplement to live flight training in the warfighter community.[1,2,3,4] Advances in technology have provided a means to create simulator environments with increased fidelity, as well as increased capabilities (e.g., distributed mission operations and live/virtual/constructive environments). Although new technologies, capable of increasing simulator fidelity, continue to become available, many questions remain unanswered about the importance of increased fidelity. For example, how should we define the trade space between increases in fidelity and the training capabilities of the simulation system, given concerns such as accessibility, training goals, and improvement costs? What skills can be trained in simulation systems versus live flight, regardless of fidelity?[6,7] What methods are available to assess the fidelity of simulation systems in order to address these concerns? What resources are available to help guide decision-making during the improvement process to maximize training benefits through improvements to the fidelity of the system? These questions are not easily answered; however, research has started to address them.

The training capabilities of a simulation system are related to the fidelity of the simulation system.[7,10] One study compared the training capabilities of a high-fidelity full simulator (the Display for Advanced Research and Technology (DART) system at AFRL, Mesa) to a low-fidelity simulation system (the Deployable Tactical Trainer (DTT) system at AFRL, Mesa). The DTT is being developed to provide a training environment capable of being located to forward operations; as such, the construction and improvement of the system are subject to numerous

---

[1] Senior Scientist, Lumir Research Institute, 1304 E. Redmon Dr., Tempe, AZ 85283, Non-member.
[2] Senior Scientist, Lumir Research Institute, 195 Bluff Ave, Grayslake, IL 60030, Non-member.
[3] Senior Research Psychologist, Human Effectiveness Directorate, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ 85212, Non-member.

American Institute of Aeronautics and Astronautics

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE **18 AUG 2008**	2. REPORT TYPE **Proceedings**	3. DATES COVERED **01-01-2007 to 31-07-2008**

4. TITLE AND SUBTITLE **Incidental Changes to Training Capabilities Due to Technological Improvements**	5a. CONTRACT NUMBER **FA8650-05-D-6502**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER **62202F**
6. AUTHOR(S) **Jason Prost; Brian Schreiber; Winston Bennett Jr**	5d. PROJECT NUMBER **1123**
	5e. TASK NUMBER **AS**
	5f. WORK UNIT NUMBER **09**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Lumir Research Institute, 301 E. Fairmont Drive,Tempe,AZ,85282**	8. PERFORMING ORGANIZATION REPORT NUMBER **AFRL; AFRL/RHA**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) **Air Force Research Laboratory/RHA, Warfighter Readiness Research Division, 6030 South Kent Street, Mesa, AZ, 85212-6061**	10. SPONSOR/MONITOR'S ACRONYM(S) **AFRL; AFRL/RHA**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) **AFRL-RH-AZ-PR-2008-0003**

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
Presented at the AIAA Modeling and Simulation Technologies Conference and Exhibit held 18-21 Aug 08 in Honolulu HI

14. ABSTRACT

During the development and enhancement of simulator systems, new technologies are incorporated to improve desired capabilities. The degree to which changes to the simulator system improve training capabilities in anticipated areas is of primary interest; however, an evaluation of benefits from newly incorporated technologies should consider the impact on training capabilities in other, less directly related areas. The current study utilized methodology developed to identify and evaluate the fidelity and deficiencies of a simulation system. The process involved utilizing Mission Essential Competencies and Subject Matter Experts to determine the training capabilities of a simulation system. The work was applied to the development of a Deployable Tactical Trainer as a method of providing feedback on the system. Data from the process were obtained throughout the development of the simulator, and the current study presents a comparison of the evaluation of the system before and after a meaningful upgrade to the technology of the simulation system. At issue is determining whether or not the changes in the training capabilities were as expected. Incidental changes to training capabilities of the system were identified and are discussed, presenting an overall evaluation of the expected versus unexpected effects on the system. Implications for the importance of building a global awareness of the simulator system are discussed. The growing dependence on simulation systems for training increases the impact of changes to the system on pilot readiness.

15. SUBJECT TERMS

Training capabilities; Technological improvements; Simulator systems; Mission Essential Competencies; System comparison; Incidental changes; Training; Pilot readiness;

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	Public Release	7	
unclassified	unclassified	unclassified			

constraints. The DART systems are full F-16 simulator training systems with full field-of-view displays and exact representations of the cockpit. Both systems operate in a Distributed Mission Operations (DMO) training environment. In the study, the training capabilities of both systems were evaluated using Subject Matter Experts (SMEs) as evaluators. Results of the study showed that the method of assessment was valid and the level of fidelity of the two systems was indeed related to the training capabilities of the system, consistent with expectations.[10]

The ability to assess the training capabilities of a simulation system and tie those capabilities to technological changes in the system (i.e., increases in fidelity) is a valuable resource for further development of the system. When considering the fidelity improvements to a simulation system, expectations are formed regarding corresponding improvements to training. Research has been conducted examining methods of identifying deficiencies of a simulation system and evaluating their effects on training.[5] System fidelity improvements may be made to directly target deficiencies thus improving the training capabilities of a system, or the changes may be undertaken with other goals in mind (e.g., enabling capabilities specifically requested or needed by a customer). Regardless of the motivating factors, the availability of efficient and effective mechanisms to evaluate the end result of fidelity improvements and determine the correspondence to expected results is lacking.

The development of a framework to evaluate the effects of fidelity improvements on training capabilities would be a valuable resource. Such a framework would be particularly important to determine those areas in which fidelity improvements did not meet expectations and those areas in which improvements resulted in unexpected, or incidental, changes in training capabilities. The ability to identify areas of training that may have been affected by improvements would be extremely important for incidental changes that led to decreases in training capabilities, as these areas may represent the introduction of negative training. The focus of the current study is to present the development of such a framework.

II. Current Study

The current study presents a process for evaluating the effects of technological changes to a simulator system. The process provides a method for identifying the degree to which changes will influence training capabilities. The goals of the current study are to 1) assess the expected improvements in the training capabilities of a system, 2) assess any actual improvements across changes, and 3) evaluate the correspondence between expected and actual improvements to a system.

The focus of evaluation in the current study was the DTT, a deployable trainer currently under development at the AFRL in Mesa, AZ. SMEs evaluated the DTT system before and after changes were made to the system's software and hardware (with the intent of increasing the training effectiveness of the system). The framework in the current study is a systematic method for evaluation of actual and expected improvements, providing a standard form of evaluation across multiple system changes.

III. Method

This study used SMEs to assess the expected versus actual improvements corresponding to various technological changes to the DTT. The following sections describe the SME evaluators, the evaluation instruments, the DTT, the changes made to the DTT, and the method of analyses used to identify differences between expected and actual improvements.

A. Evaluators

Five F-16 SMEs independently completed two evaluations of the fidelity of the DTT. The second evaluation of the DTT was conducted approximately five months after the first, following significant changes to the cockpit interface display. Evaluators participating in the current project all previously participated in similar evaluations as the one done in the current study, during a previous research on the DTT.[5,10] Evaluators were provided with detailed documentation of the system components, capabilities, and limitations, and had multiple opportunities to fly various missions in the system before completing ratings of the system.

The five SME evaluators were all male, all recently retired (4.3 years on average) F-16 pilots from the United States Air Force (USAF), with an average number of operational F-16 flight hours of 2,066 (average total flight hours of 3,814). All evaluators had been involved in the AFRL training program for at least six months prior to the first evaluation.

B. Surveys

Two surveys were utilized in the current study: The Fidelity Survey and the Anticipated Results survey. Both surveys were constructed following an established method to assess training capabilities of simulation systems.[5,10] The items on the surveys were derived from other research directed at identifying Mission Essential Competencies (MECs) critical to warfighter readiness.[6] The MEC process includes identifying knowledge, skills, and experiences necessary for a particular platform (in this study, the F-16).[6] The surveys constructed for use in the current study leveraged MEC experiences and added bold-face Emergency Procedures (EPs). Surveys included 196 total items to be rated: 70 air-to-ground (A/G) MEC experiences, 55 suppression-of-enemy air defense (SEAD) MEC experiences, 4 air-to-air (A/A) MEC experiences, and 27 bold-face EPs for the F-16.

In the Fidelity Survey, [5,10] evaluators rate each of the 196 experiences to identify the degree to which the simulation system is capable of providing the experience for training. The Fidelity survey was administered before and after significant changes had been made to the cockpit display. The response scale used was as follows:

> 0= N/A; Capability to experience does not exist
> 1= Capability to experience exists, but is very poor
> 2= Capability to experience exists, but is poor
> 3= Capability to experience exists, but is marginal
> 4= Capability to experience exists, and is good
> 5= Capability to experience exists, and is very good

The Anticipated Results Survey used the same set of 196 experiences as the Fidelity Survey. After reviewing the technological changes to the cockpit display, evaluators were asked (for each of the 196 experiences) to state whether or not the changes would decrease, maintain, or increase the training capabilities of the system.

C. Simulation System

The DTT system hardware functionality included: F-16 cockpit shell with three out-the-window 30-inch displays, the actual F-16 stick/throttle, and simulation of cockpit displays and switch functions on a high resolution 23" interactive touch screen display. Image generator was an SDS International AAcuity® PC-IG system. Brief/debrief included SmartBoard and two 50-inch displays for Head-Up Display, Radar Warning Receiver, and Multi-Function Display.

The DTT system software functionality included: The most current system uses classified Block 30 (SCU 5p) actual operational F-16 software. One database is currently installed and available. Debrief software has the ability to link and time-synchronize video recordings from multiple players. It also has the ability to network through Distributed Interactive Simulation (DIS) or High Level Architecture (HLA) standards. It has chaff/flare capability, but no ECM. Some classified weapons systems are available, such as Joint Direct Attack Munition (JDAM).

Technological changes to the system included: (1) Changes to the touch screen LCD interface for accessing switches, introducing accessibility to all cockpit display and switch functions. The earlier version only included access to essential switches. (2) Changes to the threat generator and weapons/munitions available, shifting from unclassified and generic versions to the classified, specific operational software.
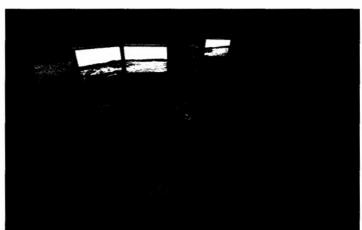

Figure 1. Example of the DTT simulation system.

D. Analytic Approach

The central focus of the current study was to provide a comparison of expected improvements to a system with actual improvements to a system. The Anticipated Results survey provided the data for expected improvements, and

the pre/post-change administration of the Fidelity survey provided the data for the actual improvements. All evaluations were based on hands-on experience with the simulation system, i.e., the SME evaluators were given the opportunity to experience all of the MECs and EPs used in the surveys.

The analytic approach used t-test comparisons of pre/post-change Fidelity survey results to determine actual improvements to the system. Additionally, the analytic approach then identified four categories of correspondence between the actual improvements to the anticipated improvements: (1) Consistency on items with no expected improvement, (2) Consistency on items with expected improvement, (3) Expected improvement with no actual improvement, and (4) No expected improvement with actual improvement. The last category represents incidental changes to the system, i.e., changes that were not anticipated.

IV. Results

Survey responses from SME evaluators were compiled in an MS Excel spreadsheet for analysis. An inter-rater reliability score and an agreement score were computed for the pre- and post-change evaluations of training capabilities (DTT pre-change Fidelity Survey and DTT post-change Fidelity Survey). Additionally, an agreement score was computed for the Anticipated Results survey. The Anticipated Results survey responses were dichotomous; therefore, inter-rater reliability was not computed. Inter-rater reliability calculations were computed by correlating scores for each pair of raters and taking the average across all correlations.[11] The agreement calculations were made after converting rating responses into a discrete decision choice of either N/A (i.e., 0 ratings) or some degree of applicability (i.e., any rating of 1 or higher). Table 1 presents the inter-rater reliability and agreement scores.

Table 1. Inter-rater reliability and agreement scores

	DTT pre-chg	DTT post-chg	Anticipated Results
Inter-rater Reliability	.56	.75	N/A
Inter-rater Agreement	.70	.86	.88

To identify the MEC experiences and EPs for which system changes influenced training capabilities, SME evaluations from pre-change and post-change were analyzed. Paired sample t-tests were conducted to identify which of the MEC experiences and EPs were rated as having significantly improved from pre- to post-change. The paired sample t-tests were conducted using an alpha level of 0.1, due to the low sample size and low cost of Type One errors in this research. Of the 196 MEC experiences and EPs, 22 experiences were rated significantly different from pre- to post-change.. The average increase in the ratings from pre- to post-change evaluations was 0.89. In the current study the only changes in the training capabilities of the system could all be characterized as improvements, there were no negative changes in training capabilities. These results support the ability of the approach to identify improvements in the training capabilities of a simulation system using a SME evaluator approach.

The next step in the analyses was to identify the correspondence between anticipated improvements and actual improvements to the training capabilities. Table 2 presents the frequency and proportion of experiences falling into each of four categories: consistency on items with no expected improvement, consistency on items with expected improvement, expected improvement with no actual improvement, and no expected improvement with actual improvement.

Table 2. Correspondence of anticipated and actual improvements.

Anticipated Results	Actual Results	f	%
Improvement	Sig. Change	20	10
	No Sig. Change	24	12
No Improvement	Sig. Change	2	1
	No Sig. Change	150	77

V. Conclusion

The findings of the current study suggest that the method proposed to evaluate the influence of technological changes to a simulation system on the training capabilities is both efficient and sensitive for this domain. It was found that SME pilots were extremely accurate in identifying training experiences that would not benefit from the technological changes to a system. Across 196 MEC experiences and EPs, pilots anticipated that 152 of those experiences would realize no benefits from upcoming changes to the system. An evaluation of the training capabilities after the system changes were made revealed no actual benefit for 150 of the experiences. This finding

reflects a 98.7% accuracy rate by SMEs in identifying experiences that would not be affected by changes to the technology.

Analyses of the experiences for which SMEs did anticipate improvement revealed less conclusive evidence of accurate identification. SME evaluators anticipated that 44 experiences would improve with the technological changes to the system. Of those 44 experiences, only 20 significantly improved from pre- to post-change evaluation. Possible explanations for these results include statistical power issues (due to small sample size) or the possibility that the technological changes to the system did not have the anticipated result. Further investigation found that the training capability rating for 19 of the 24 experiences increased from pre- to post-change evaluation, the remaining 5 showing no change across evaluations. This finding suggests that low statistical power may be the more reasonable explanation of the result.

The technological system changes in the current study were undertaken to increase the fidelity of the DTT system by improving the threat generation system, available weapons, and providing a touch screen display that allowed access to all cockpit displays and switches. The purpose of these changes was to increase training capabilities of the system, although the fidelity of those changes did not represent a perfect mapping to the actual cockpit. Due to the constraints of a deployable system, access to displays and switches was provided through a single touch screen, so those displays and switches could be accessed despite the shift in the switchology of the system. The added functionality of the DTT system was also undertaken as an intermediate step to facilitate future changes to the system. Future changes require full switch and display accessibility (i.e., the introduction of Block 40 software capabilities). Therefore, the post-change evaluation of the system may not reflect the full degree of change. Some of the intended benefits may additional changes to the system.

The experiences that were found to have improved significantly in the current study were primarily those pertaining to threats and weapons. For example, reacting to realistic type threat (e.g., SAM, AAA) at low altitude, 1:2 Force ratio (e.g., 4 v 8), 1:3+ Force ratio (e.g., 4 v 12+), and OCA escort missions all improved. The force ratio experiences are directly related to the changes to the system increasing the fidelity of the threat and weapons systems. Force ratio experiences pertain to engagements in which the ratio of friendly to threat aircraft is uneven in the favor of the enemy. The experiences that SME evaluators anticipated would improve, but showed no actual improvement pre- to post-change were primarily those involving changes to the touch screen. For example, targeting and sorting responsibilities, operations against a threat using chaff/flare, emergency or selective jettison action, and exercise jettison criteria were identified. The pattern of results suggested that changes to the threat generator and the weapons systems may have produced larger effects than the change to the touch screen display.

The two experiences that significantly improved despite SMEs not anticipating improvement were fatigue/time on task (e.g., long-range force employment) and operations against a threat that employs chaff/flare. These two experiences represent what might be considered incidental changes to the system. These changes may not have been anticipated for very different reasons. The ability of the threat generator to incorporate threats that employ chaff/flare was not known to the SME evaluators at the time of assessment. When considering future improvements, it will be important to fully inform SME evaluators of not only the current capabilities, but also the potential changes to those capabilities that are currently being worked on (such as a threat generator being updated with newer capabilities). The second finding was likely an unanticipated result of changes to the weapons system and display. Both of these changes provided a means for pilots to use all of the tools normally at their disposal, thus mitigating some of the fatigue and frustration previously experienced in the more limited simulator.

The current study provides a framework to investigate the anticipated improvements of simulator system changes and to compare those expectations to the actual impact on training capabilities. The results of this study suggest that the framework is an efficient and valid approach to investigating improvements to system capabilities as a result of technological changes. Further, the framework can be used to compare anticipated improvements with actual training improvements of a simulation system.

Although the results were consistent with the intention of the assessment framework, three possible limitations of the methodology arise. The following list of limitations may not be exhaustive, but represent the most pressing issues in the opinion of the authors.

First, the system evaluations in the current study showed very strong agreement among SME evaluators regarding anticipated improvements in training capabilities due to technological system changes; however, the inter-rater reliability was lower than desired for the first assessment of the DTT system. Due to the small sample size in the current study, the variability in responses across raters has a large effect on the sensitivity of statistical tests in

identifying changes in ratings. A possible solution would be to increase the number of SME evaluators in order to increase the power of the statistical tests and/or to work with SME evaluators to ensure that they are completing assessments with the same sets of assumptions (i.e., try to reduce variability across raters).

The second potential limitation in the current study involves the use of a categorical variable to measure the anticipated improvements to a system. The categorical variable does not allow for measurement of the degree of anticipated change resulting from the improvement. In the current study, there were a number of experiences that the SME evaluators anticipated would improve due to changes in technology; yet, the evaluation of the training capabilities did not support those expectations statistically. Looking at the average change from pre- to post-change evaluations (an increase of .48), there is evidence that evaluators indicated some improvement to the system across those experiences and in the appropriate direction. A solution to this limitation would be to allow SME evaluators the ability to assess anticipated improvements using a scale with more discriminability.

A final limitation of the current study is the lack of strong findings for incidental changes to the system. Changes made to the touch screen to provide access to all cockpit displays and switches was of particular interest in the study. The changes required pilots to adjust their switchology and adapt to a cockpit environment. This low-fidelity solution to the problem of accessibility to ship functions may have introduced negative training or resulted in attenuation of training capabilities. The findings of the current study did not suggest any negative training introduced by the change; indeed, the only incidental changes to the system positively influenced training capabilities. The limitation presented here is the inability to know if the use of SME evaluators and this framework is sensitive to negative changes in training capabilities brought about by system changes. There are no suggestions for how to resolve this issue, although the application of this framework to the system across upcoming technological changes will provide another opportunity to use the framework.

The current framework is an extremely efficient and effective mechanism for evaluating the effects of changes to system technology. The application of this framework during the development of a system, particularly during a spiral approach to development, is a powerful feedback tool to inform changes. Future application of the framework will hopefully establish the ability of the process to identify incidental changes to a system that are negative in nature. The framework provides a valid means for investigating the degree to which changes have improved training capabilities and incidental changes that are positive in nature.

Acknowledgments

References

[1]Bell, H. H. & Waag, W. L. Evaluating the effectiveness of flight simulators for training combat skills: A review. The International Journal of Aviation Psychology. 8 (3), pp. 223-242, 1998.

[2]Gehr, S. E., Schreiber, B. T., & Bennett, W. Jr. Within-Simulator Training Effectiveness Evaluation. Interservice/Industry Training, Simulation and Education Conference (I/ITSEC) Proceedings. Orlando, FL: National Security Industrial Association, 2004.

[3]Schreiber, B. T. & Bennett, W. Jr. Distributed Mission Operations within simulator training effectiveness baseline study: Summary report. (AFRL-HE-AZ-TR2006-0015-Vol I). Air Force Research Laboratory, Warfighter Readiness Research Division, Mesa AZ, 2006.

[4]Schreiber, B. T., Rowe, L. J., & Bennett, W. Jr. Distributed Mission Operations within-simulator training effectiveness baseline study: Participant utility and effectiveness opinions and ratings. (AFRL-HE-AZ-TR-2006-0015-Vol IV). Mesa AZ: Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Readiness Research Division, Mesa AZ, 2006.

[5]Prost, J.H., Schreiber, B. T., & Bennett, W. Jr. Identification and Evaluation of Simulator System Deficiencies. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) Proceedings. Orlando, FL: National Security Industrial Association, 2007.

[6]Colegrove, C. M. & Alliger, G. M. Mission Essential Competencies: Defining Combat Mission Readiness in a Novel Way. Paper presented at the NATO RTO Studies, Analysis and Simulation (SAS) Panel Symposium. Brussels, Belgium, April 2002.

[7]Allen, J., Buffardi, L., & Hays, R. The Relationship of Simulator Fidelity to Task and Performance Variables (ADA238941). Fairfax, VA: George Mason University. 1991.

[8]Dennis, K. A. & Harris, D. Computer-based simulation as an adjunct to ab initio flight training. International Journal of Aviation Psychology, 8, 261-276, 1998.

[9]Schank, J. F., Thie, H. J, Graf, C. M., Beel, J., & Sollinger, J. Finding the right balance: Simulator and Live Training for Navy Units. Santa Monica: RAND Corporation. 2002.

[10]Schreiber, B. T., Gehr, S. E., & Bennett, W. Jr. Fidelity Trade-Offs for Deployable Training and Rehearsal. Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) Proceedings. Orlando, FL: National Security Industrial Association. 2006.

[11]Glass, G. V. & Hopkins, K. D. (1995). Statistical Methods in Education and Psychology, 3rd Ed. Needham Heights, MA: Allyn and Bacon.